Journal of Multivariate Analysis **67**, 297–305 (1998)
Article No. MV981771

A Generalization of Rao's Covariance Structure with Applications to Several Linear Models

Hiroshi Kurata*

Yamaguchi University, Yamaguchi, Japan

Received April 22, 1997; revised May 5, 1998

This paper presents a generalization of Rao's covariance structure. In a general linear regression model, we classify the error covariance structure into several categories and investigate the efficiency of the ordinary least squares estimator (OLSE)

OLSE and the GME. Hence our classification includes Rao's covariance structure. The results are applied to models with special structures: a general multivariate analysis of variance model, a seemingly unrelated regression model, and a serial correlation model.

Key words and phrases: Gauss–Markov estimator; ordinary least squares estimator; Rao's covariance structure; seemingly unrelated regression model; general multivariate analysis of variance model.

1. INTRODUCTION

In a general linear regression model

$$y = X\beta + \varepsilon \qquad \text{with} \quad E\varepsilon = 0 \qquad \text{and} \qquad E\varepsilon\varepsilon' = \sigma^2 \Omega, \tag{1.1}$$

where y: $n \times 1$, X: $n \times k$, rank $X = k$ and $\Omega \in S^+(n)$, we shall investigate the efficiency of least squares estimators of the coefficient vector β via classifying the covariance structure of Ω. Here $S^+(n)$ denotes the set of $n \times n$ positive definite matrices.

Provided that Ω is known, the Gauss–Markov estimator (GME)

$$\hat{\beta}(\Omega) = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$$

* The author would like to express his sincere gratitude to Professors Takeaki Kariya and Takesi Hayakawa of Hitotsubashi University for their valuable suggestions and encouragements. He deeply thanks the editors and an anonymous referee for quite significant comments on the earlier version.

0047-259X/98 $25.00

is the best linear unbiased estimator of β. In most cases, however, Ω is unknown and we often estimate β by the ordinary least squares estimator (OLSE) which is defined as

$$\hat{\beta}(I_n) = (X'X)^{-1} X'y.$$

Throughout this paper, we call $\hat{\beta}(\Omega)$ the GME even when Ω is unknown.

By the Gauss–Markov theorem, $\mathrm{Cov}\,\hat{\beta}(I_n) \geqq \mathrm{Cov}\,\hat{\beta}(\Omega)$ holds for any X and Ω, where the inequality should be understood in terms of non-negative definiteness. The efficiency loss $\mathrm{Cov}\,\hat{\beta}(I_n) - \mathrm{Cov}\,\hat{\beta}(\Omega)$ depends on the relation between the structures of X and Ω. Many papers have investigated the efficiency of the OLSE relative to the GME using various one-dimensional measures. For example, Bloomfield and Watson (1975) and Knott (1975) used the ratio of the generalized variances $\eta \equiv |\mathrm{Cov}\,\hat{\beta}(\Omega)|/|\mathrm{Cov}\,\hat{\beta}(I_n)|$ and obtained its bounds in terms of the latent roots of Ω. The η as well as other one-dimensional measures, however, does not fully reflect the structural relation between X and Ω, or equivalently, the relation between $L(X)$ and $L(\Omega X)$, where $L(X)$ is the column space of X. That is, $\eta = 1$ does not necessarily imply that

$$\hat{\beta}(I_n) \equiv \hat{\beta}(\Omega), \tag{1.2}$$

where $\equiv$ means that the equality holds for all $y \in R^n$. Further, such measures quite often depend on the unknown Ω.

A lot of researches are found on necessary and sufficient conditions for (1.2). Among them, Rao (1967) and Geisser (1970) proved that (1.2) holds for given X if and only if Ω is written as

$$\Omega = (X, Z) \begin{pmatrix} \Gamma & 0 \\ 0 & \Delta \end{pmatrix} \begin{pmatrix} X' \\ Z' \end{pmatrix} \tag{1.3}$$

for some $\Gamma \in S^+(k)$ and $\Delta \in S^+(n-k)$, where Z is any $n \times (n-k)$ matrix such that $X'Z = 0$ and $\mathrm{rank}\, Z = n-k$. The structure (1.3) of Ω is called Rao's covariance structure. Other characterizations of (1.2) are, for example, that $L(X)$ is spanned by some k latent vectors of Ω (Zyskind, 1967) and that $L(X)$ is Ω-invariant (Kruskal, 1968). Further, Zyskind (1969) gave general conditions on the structure of Ω for which (1.2) holds for all X such that $L(X) \supset L(U)$ where U is a fixed matrix. This result was generalized by Mathew (1983) to the case where Ω is incorrectly specified.

In most cases, Ω may not be expressed as (1.3) and hence the OLSE may deviate from the GME. The degree of deviation between the two estimators depends on the structure of Ω. Therefore we need to provide a more detailed classification of the covariance structure of Ω in order to investigate how much the degree of deviation is. In this paper, as a classification criterion, we adopt the rank of the covariance matrix of the difference between the OLSE and the GME:

$$\operatorname{rank} \operatorname{Cov}(\hat{\beta}(I_n) - \hat{\beta}(\Omega)) \equiv v(\Omega, X). \tag{1.4}$$

The criterion $v(\Omega, X)$ is an integer valued function satisfying $0 \leqq v(\Omega, X) \leqq \min(k, n-k)$. Clearly, the covariance structure corresponding to $v(\Omega, X) = 0$ is (1.3). But the structures corresponding to each value of $0 < v(\Omega, X) \leqq \min(k, n-k)$ have been left unknown. This problem is considered in Section 2. The result shall be extended to the problem of simultaneous estimation of β and σ^2. $v(\Omega, X)$ equals the dimension of the linear subspace in which the difference of two estimators realizes. In Section 3, it is proved that $v(\Omega, X)$ does not depend on Ω under several linear models.

2. CLASSIFICATION ACCORDING TO $v(\Omega, X)$

Under the model (1.1), we have

$$\begin{aligned}\operatorname{Cov}(\hat{\beta}(I_n) - \hat{\beta}(\Omega)) &= \sigma^2\{(X'X)^{-1} X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1}\},\\ &\equiv \sigma^2 \Psi(\Omega, X).\end{aligned} \tag{2.1}$$

Since $\Psi(\Omega, XG) = G^{-1}\Psi(\Omega, X)\, G^{-1'}$ holds for any $k \times k$ non-singular matrix G, $v(\Omega, X)$ depends on X only through $L(X)$. Any $\Omega \in S^+(n)$ is expressed as

$$\Omega = (X, Z)\begin{pmatrix}\Gamma & \Xi\\ \Xi' & \Delta\end{pmatrix}\begin{pmatrix}X'\\ Z'\end{pmatrix} \tag{2.2}$$

for some $\Gamma \in S^+(k)$, $\Delta \in S^+(n-k)$ and Ξ: $k \times (n-k)$, where Z is defined in (1.3). By using $(X, Z)^{-1} = [X(X'X)^{-1}, Z(Z'Z)^{-1}]'$, the inverse matrix of Ω is calculated as

$$\Omega^{-1} = (X, Z)\begin{pmatrix}\Gamma_0 & \Xi_0\\ \Xi_0' & \Delta_0\end{pmatrix}\begin{pmatrix}X'\\ Z'\end{pmatrix}, \tag{2.3}$$

where

$$\Gamma_0 = (X'X)^{-1}(\Gamma - \Xi\Delta^{-1}\Xi')^{-1}(X'X)^{-1} \in S^+(k),$$
$$\Delta_0 = (Z'Z)^{-1}(\Delta - \Xi'\Gamma^{-1}\Xi)^{-1}(Z'Z)^{-1} \in S^+(n-k)$$

and

$$\Xi_0 = -(X'X)^{-1}\Gamma^{-1}\Xi(\Delta - \Xi'\Gamma^{-1}\Xi)^{-1}(Z'Z)^{-1}\colon k \times (n-k).$$

Substituting the right hand sides of (2.2) and (2.3) into $\Psi(\Omega, X)$ yields $\Psi(\Omega, X) = \Xi\Delta^{-1}\Xi'$. Hence we have the following result.

THEOREM 2.1. *Under the model* (1.1),

$$v(\Omega, X) = \text{rank } \Xi$$

holds. That is, $v(\Omega, X) = v$ *holds for an integer* v *satisfying* $0 \leqq v \leqq \min(k, n-k)$ *if and only if* Ω *is written as*

$$\Omega = (X, Z)\begin{pmatrix} \Gamma & \Xi \\ \Xi' & \Delta \end{pmatrix}\begin{pmatrix} X' \\ Z' \end{pmatrix} \tag{2.4}$$

for some $\Gamma \in S^+(k)$, $\Delta \in S^+(n-k)$ *and* Ξ: $k \times (n-k)$ *such that* rank $\Xi = v$.

Thus the structure of Ω is classified according to $v(\Omega, X)$ into $\min(k, n-k) + 1$ categories. In the sequel, we shall denote the covariance structure (2.4) by $CS(v)$. Clearly $CS(0)$ is equivalent to Rao's covariance structure. The following formula is an easy consequence of (2.2) and (2.3):

$$v(\Omega, X) = v(\Omega^{-1}, X) = \text{rank } X'\Omega Z.$$

This value equals the dimension of the linear subspace in which the difference between the OLSE and the GME realizes. It reflects the degree of departure from Rao's covariance structure.

Next we shall consider the simultaneous estimation problem of β and σ^2. We define the GM-type estimator $s^2(\Omega)$ and the OLS-type estimator $s^2(I_n)$ of σ^2 as

$$s^2(\Omega) = (y - X\hat{\beta}(\Omega))'\, \Omega^{-1}(y - X\hat{\beta}(\Omega))/(n-k),$$

and

$$s^2(I_n) = (y - X\hat{\beta}(I_n))'\,(y - X\hat{\beta}(I_n))/(n-k),$$

respectively. The following results are given by Kariya (1980).

(K-1) $s^2(\Omega) \equiv s^2(I_n)$ *holds if and only if* Ω *is written as*

$$\Omega = N + \Lambda - N\Lambda N \tag{2.5}$$

for some $n \times n$ *symmetric matrix* Λ, *where* $N = Z(Z'Z)^{-1}Z'$.

(K-2) *$\hat{\beta}(\Omega) \equiv \hat{\beta}(I_n)$ and $s^2(\Omega) \equiv s^2(I_n)$ hold if and only if Ω is written as*

$$\Omega = X\Gamma X' + N \tag{2.6}$$

for some $\Gamma \in S^+(k)$.

It is noted that Kariya's covariance structure (2.6) is equivalent to $CS(0)$ in (2.4) with $\Delta = (Z'Z)^{-1}$. The result below is a natural extension of Kariya (1980).

THEOREM 2.2

$$v(\Omega, X) = v \qquad \text{and} \qquad s^2(\Omega) \equiv s^2(I_n) \tag{2.7}$$

holds for an integer v satisfying $0 \leqq v \leqq \min(k, n-k)$ if and only if Ω is of the form $CS(v)$ with $\Delta = (Z'Z)^{-1}$, or equivalently, Ω is written as

$$\Omega = (X, Z) \begin{pmatrix} \Gamma & \Xi \\ \Xi' & (Z'Z)^{-1} \end{pmatrix} \begin{pmatrix} X' \\ Z' \end{pmatrix} \tag{2.8}$$

for some $\Gamma \in S^+(k)$ and Ξ: $k \times (n-k)$ such that rank $\Xi = v$.

Proof. [Necessity] Suppose that (2.7) holds. Then by Theorem 2.1, Ω is written as

$$\Omega = X\Gamma X' + X\Xi Z' + Z\Xi' X' + Z\Delta Z'$$

for some $\Gamma \in S^+(k)$, $\Delta \in S^+(n-k)$ and Ξ: $k \times (n-k)$ such that rank $\Xi = v$. It remains to show that $\Delta = (Z'Z)^{-1}$. From (K-1), Ω is also written as (2.5) for some symmetric matrix Λ. Hence we get the equality

$$X\Gamma X' + X\Xi Z' + Z\Xi' X' + Z\Delta Z' = N + \Lambda - N\Lambda N.$$

Premultiplying by $(Z'Z)^{-1} Z'$ and postmultiplying by $Z(Z'Z)^{-1}$ yield $\Delta = (Z'Z)^{-1}$. This proves the necessity.

[Sufficiency] Suppose that Ω is of the form (2.8). Then it is clear from Theorem 2.1 that $v(\Omega, X) = v$. To prove that $s^2(I_n) \equiv s^2(\Omega)$, let $\Lambda = X\Gamma X' + X\Xi Z' + Z\Xi' X'$. Since $\Lambda = \Lambda'$ and $N\Lambda N = 0$ hold, Ω can be written as (2.5), proving the result. ∎

The results above are applicable to the model with linear restrictions via reducing the model to a model without restrictions. See, for example, Rao (1973, Section 4a.9). (This point is due to the referee.)

3. APPLICATIONS

In this section, we briefly consider the cases where the regression matrix X and the error covariance matrix $\sigma^2\Omega$ of the model (1.1) have some special structure. In such cases Ω is given by a function of an unknown vector θ: $\Omega=\Omega(\theta)$. Models considered here are a general multivariate analysis of variance (GMANOVA) model, a seemingly unrelated regression (SUR) model and a serial correlation model. It is proved that under the last two models, $v(\Omega(\theta), X)$ does not depend on the unknown θ.

EXAMPLE 1. GMANOVA model.

The GMANOVA model considered here is given as

$$Y=X_1BX_2'+E \qquad \text{with} \quad \mathrm{Cov}(\mathrm{vec}(E'))=I_n\otimes\sigma^2\Sigma \tag{3.1}$$

where Y: $n\times p$, $X_1: n\times k$, rank $X_1=k$, X_2: $p\times m$ and rank $X_2=m$ and $\sigma^2\Sigma\in S^+(p)$. The GME $\hat{B}(\Sigma)$ and the OLSE $\hat{B}(I_p)$ of the coefficient matrix B are given by

$$\hat{B}(\Sigma)=(X_1'X_1)^{-1}X_1'Y\Sigma^{-1}X_2(X_2'\Sigma^{-1}X_2)^{-1}$$

and

$$\hat{B}(I_p)=(X_1'X_1)^{-1}X_1'YX_2(X_2'X_2)^{-1}$$

respectively. Since $\mathrm{Cov}(\hat{B}(I_p)-\hat{B}(\Sigma))=(X_1'X_1)^{-1}\otimes\sigma^2\Psi(\Sigma, X_2)$, its rank is given by

$$\text{rank Cov}(\hat{B}(I_p)-\hat{B}(\Sigma))=k\times v(\Sigma, X_2).$$

Let Z_2 be any $p\times(p-m)$ matrix satisfying $X_2'Z_2=0$ and rank $Z_2=p-m$. Then applying Theorem 2.1 yields the following result.

COROLLARY 3.1.

$$\text{rank Cov}(\hat{B}(I_p)-\hat{B}(\Sigma))=k\times v$$

holds for an integer $0\leqq v\leqq\min(m, p-m)$ *if and only if* Σ *is of the form* $CS(v)$, *or equivalently*, Σ *is expressed as*

$$\Sigma=(X_2, Z_2)\begin{pmatrix}\Gamma & \Xi\\ \Xi' & \Delta\end{pmatrix}\begin{pmatrix}X_2'\\ Z_2'\end{pmatrix} \tag{3.2}$$

for some $\Gamma\in S^+(m)$, $\Delta\in S^+(p-m)$ *and* Ξ: $m\times(p-m)$ *such that* rank $\Xi=v$.

The case where $v=0$ in (3.2) is given by Kariya (1985).

The GM-type estimator $\hat{\sigma}^2(\Sigma)$ of σ^2 is defined as

$$\hat{\sigma}^2(\Sigma)=\mathrm{tr}[\Sigma^{-1}(Y-X_1\hat{B}(\Sigma)X_2')'(Y-X_1\hat{B}(\Sigma)X_2')]/(np-km),$$

and we call $\hat{\sigma}^2(I_p)$ the OLS-type estimator of σ^2. Kariya (1985) derived the following two results as an application of (K-1) and (K-2).

(K-3) *A necessary and sufficient condition for which* $\hat{\sigma}^2(I_p)\equiv\hat{\sigma}^2(\Sigma)$ *is that* Σ *is written as*

$$\Sigma=I_p+H-N_2HN_2$$

for some $p\times p$ *symmetric matrix* H, *where* $N_2=Z_2(Z_2'Z_2)^{-1}Z_2'$.

(K-4) *A necessary and sufficient condition for which* $\hat{\sigma}^2(I_p)\equiv\hat{\sigma}^2(\Sigma)$ *and* $\hat{B}(I_p)\equiv\hat{B}(\Sigma)$ *is that* Σ *is written as*

$$\Sigma=X_2\Gamma X_2'+N_2$$

for some $\Gamma\in S^+(m)$. *That is,* Σ *is written as* $CS(0)$ *with* $\Delta=(Z_2'Z_2)^{-1}$.

As an extension of Kariya's results, we obtain the following corollary.

Corollary 3.2.

$$\mathrm{rank}\,\mathrm{Cov}(\hat{B}(I_p)-\hat{B}(\Sigma))=k\times v \qquad \text{and} \qquad \hat{\sigma}^2(\Sigma)\equiv\hat{\sigma}^2(I_p)$$

hold for an integer v *satisfying* $0\leqq v\leqq\min(m,p-m)$ *if and only if* Σ *is written as* $CS(v)$ *with* $\Delta=(Z_2'Z_2)^{-1}$.

Example 2. SUR model.

The SUR model considered here is defined as the model (1.1) with the structure

$$\begin{aligned} &y=(y_1',y_2')', && X=\mathrm{diag}\{X_1,X_2\} \\ &\beta=(\beta_1',\beta_2')', && \varepsilon=(\varepsilon_1',\varepsilon_2')', \\ &\Omega=\Sigma\otimes I_m, && \Sigma=(\sigma_{ij})\in S^+(2) \quad \text{and} \quad \sigma^2=1, \end{aligned} \tag{3.3}$$

where y_i: $m\times 1$, X_i: $m\times k_i$, rank $X_i=k_i$, ε_i: $m\times 1$, β_i: $k_i\times 1$, $n=2m$, $k=k_1+k_2$ and diag$\{X_1,X_2\}$ denotes block diagonal matrix. Kariya (1981) proved that $v(\Sigma\otimes I_m, X)=0$ holds if and only if $L(X_1)=L(X_2)$. Let Z_i be any $m\times(m-k_j)$ matrix such that $X_j'Z_i=0$ and rank $Z_j=m-k_j$, and take $Z = \mathrm{diag}\{Z_1, Z_2\}$. Corollary 3.3 below is a direct consequence of the following equality

$$X'(\Sigma \otimes I_m)\, Z = \sigma_{12} \begin{pmatrix} 0 & X_1' Z_2 \\ X_2' Z_1 & 0 \end{pmatrix}.$$

COROLLARY 3.4. *If* $\sigma_{12} \neq 0$, *then*

$$v(\Sigma \otimes I_m, X) = \mathrm{rank}\ X_1' Z_2 + \mathrm{rank}\ X_2' Z_1. \tag{3.4}$$

Note that the right hand side of (3.4) is free from the unknown Σ.

In the case where $L(X_1) \subset L(X_2)$ ($k_1 \leqq k_2$) (Revankar (1974)), it follows from (3.4) that

$$v(\Sigma \otimes I_m, X) = 0 + (k_2 - k_1) = k_2 - k_1.$$

As was proved by Kariya (1981), $L(X_1) \subset L(X_2)$ is a necessary and sufficient condition so that the identity between the OLSE and the GME of the coefficient vector β_1 of the first equation holds.

In the case where $X_1' X_2 = 0$ (Zellner, 1963), v is calculated as

$$v(\Sigma \otimes I_m, X) = k_1 + k_2 = k.$$

EXAMPLE 3. Serial correlation models.

We treat the model (1.1) with the structure

$$\Omega(\theta)^{-1} = I_n + \theta A \qquad \text{with} \quad \theta \in R^1 - \{0\}$$

where A is a known symmetric matrix. Such models include an intra-class correlation model, a first order autoregressive error model, a circularly correlated model, and a 2-equation heteroscedastic model. In these models

$$v(\Omega(\theta), X) = \mathrm{rank}\ X' A Z$$

holds and is free from θ, where Z is any $n \times (n-k)$ matrix satisfying $X'Z = 0$ and rank $Z = n - k$. Some related topics are found in Usami and Toyooka (1997).

REFERENCES

P. Bloomfield and G. S. Watson, The inefficiency of least squares, *Biometrika* **62** (1975), 121–128.

S. Geisser, Bayesian analysis of growth curve, *Sankhya* (*A*) **32** (1970), 53–64.

T. Kariya, A note on a condition for equality of sample variances in a linear model, *J. Amer. Statist. Assoc.* **75** (1980), 701–703.

T. Kariya, Test for the independence between two seemingly unrelated regression equations, *Ann. Statist.* **9** (1981), 381–390.

T. Kariya, "Testing in the Multivariate General Linear Model," Kinokuniya, Tokyo, 1985.

M. Knott, On the minimum efficiency of least squares, *Biometrika* **62** (1975), 129–132.

W. Kruskal, When are gauss–Markov and least squares estimators identical? A coordinate free approach, *Ann. Math. Statist.* **39** (1968), 70–75.

T. Mathew, Linear estimation with an incorrect dispersion matrix in linear models with a common linear part, *J. Amer. Statist. Assoc.* **78** (1983), 468–471.

C. R. Rao, Least squares theory using an estimated dispersion matrix and its application to measurement of signals, *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** (1967), 355–372.

C. R. Rao, "Linear Statistical Inference and Its Applications," 2nd ed., Wiley, New York, 1973.

N. S. Revankar, Some finite sample results in the context of two seemingly unrelated regression equations, *J. Amer. Statist. Assoc.* **69** (1974), 187–190.

Y. Usami and Y. Toyooka, On the degeneracy of the distribution of a GLSE in a regression with circularly distributed error, *Math. Japonica* **45** (1997), 423–432.

A. Zellner, Estimators for seemingly unrelated regression equations: some exact finite sample results, *J. Amer. Statist. Assoc.* **58** (1963), 977–992.

G. Zyskind, On canonical forms, nonnegative covariance matrices, and best and simple least squares estimators in linear models, *Ann. Statist.* **38** (1967), 1092–1110.

G. Zyskind, Parametric augmentations and error structures under which certain least squares and analysis of variance procedure are also best, *J. Amer. Statist. Assoc.* **64** (1969), 1353–1368.